EECS 553 FALL 2022 SYLLABUS, MACHINE LEARNING - ECE*

The goal of machine learning is to develop computer algorithms that can learn from data or past experience to make accurate, useful, and fair predictions on new unseen data. In the past few decades, machine learning has grown tremendously, and it has made major impacts in many real-world applications. This course will give a graduate-level introduction of machine learning and provide mathematical/statistical foundations of machine learning, mathematical derivation and implementation of the algorithms, and discussion of their applications. On assignments, students will work through mathematical derivations applying principles learned in class, implement machine learning algorithms in Python, and apply those algorithms to data sets spanning a variety of applications. In small groups, students will also participate in the Fall 2022 Machine Learning Reproducibility challenge, which involves reproducing the results and assessing the conclusions of a paper published in a major machine learning conference.

Instructor: Professor Laura Balzano, girasole@umich.edu

Course time: Tuesdays and Thursdays, 1:30-2:50pm US Eastern time

Course location: GGBL 2505. All lectures will be recorded.

GSI(s): Zongyu Li - zonyul@umich.edu, Sudeep Katakol - sudeepk@umich.edu
Zhe Du - zhedu@umich.edu, Kyle Gilman - kgilman@umich.edu, Haonan Zhu - haonan@umich.edu

Advisory Pre-requisites: EECS 501 and 551 (or related courses). A curiosity about the mathematical models and representations behind machine learning algorithms. Patience with abstraction and with the required effort to translate mathematical models into algorithms and applications.

Required Textbook: None. All homeworks and quizzes will be based on lecture notes and slides.
Additional useful texts are listed on Canvas>Files>MLTexts.pdf
Office hours: These may change if we lose GSIs, due to the fact that 553 enrollment is not full

Zhe Du	Monday 4 - 5	(EECS 3312)
Zongyu Li	Tuesday 10 - 11	(zoom)
Haonan Zhu	Tuesday 3 - 4	(zoom)
Kyle Gilman	Tuesday 4 - 5	(EECS 4419)
Sudeep Katakol	Wednesday 9 - 10	(zoom)
Zongyu Li	Thursday 10 - 11	(zoom)
Laura Balzano	Friday 11 - 12	(zoom)
Sudeep Katakol	Saturday 11 - 12	(zoom)

Communication and Course Admin: The course will be run through Canvas. All materials will be posted there. Announcements in Canvas will be the main venue for communication from the teaching staff to students. Lecture recordings will be available through Canvas. Less critical messages from teaching staff may also be sent through Piazza. Piazza through Canvas will be used for class discussions. Gradescope will be linked there for homework submission (entry code J37GNJ). iClicker will be used for in-class participation.

Grade breakdown:
Quiz 20%, HW 30%, ML Reproducibility Project 35%, Floater 5% takes on your highest grade in the above three categories
Clicker questions 5%, Piazza participation 5%

Quizzes: There will be one evening quiz during the semester **on Tuesday, 11 October, 7-9pm**

*This is an official version as of September 15, 2022.

You will be allowed two note sheets (8.5 × 11", both sides), but it is otherwise closed-book. The quiz will be mostly multiple choice. You will have access to a practice quiz one week before the quiz. There will be options for making up points on the quiz if you do poorly, announced approximately Tuesday, 18 October.

If you have other mandatory obligations that conflict with the quiz date, please inform the entire teaching staff with a private Piazza message using subject "EECS 553 - quiz conflict" as soon as possible and two weeks before the quiz at the latest. Those students will take the quiz during class time on the same day. If you have SSD accommodations for exams, please send your documentation to Professor Balzano with the subject line "EECS 553 - SSD" at the start of the semester, two weeks before the first quiz at the latest. If you send the documentation through the "Accommodate" system, please also send the professor an email with the above subject line, which helps me quickly find you when the quiz time comes.

If you get sick or test positive for COVID-19 before the quiz, please inform the entire teaching staff with a private Piazza message as soon as possible with subject line "EECS 553 - illness". If it is within 3 hours of the start time, please send an email to Professor Balzano and the rest of the teaching staff with subject line "EECS 553 - illness". Either way you must send your message before the quiz. We will quickly arrange an alternative option for you.

Homework: Homework will be due on the following schedule (which is subject to change), at 11:59pm unless otherwise noted:
HW 1 - Sept 6 - probability and linear algebra background
HW 2 - Sept 20
HW 3 - Oct 4
HW 4 - Oct 25
HW 5 - Nov 15

Late homework will be accepted until 48 hours later with a penalty of 25% per day (any portion of 24 hours). No homework will be accepted beyond that point. Your lowest homework grade will be dropped. If you find that you might miss 2 or more homeworks due to health or personal reasons, please inform the entire teaching staff with a private Piazza message with subject line "EECS 553 - missed homework" within two days of the second missed homework to discuss your circumstances. Until 48 hours have passed after each homework deadline, we will not post/discuss solutions on Canvas, and please do not discuss specific solutions on Piazza.

Homeworks will involve both pen/paper exercises and programming in Python. Every homework will also have the students submit at least one multiple choice question on the current subject matter along with its solution (see more below). Solutions or proposed questions in languages other than Python are not acceptable. Prior experience with Python is not assumed, although experience with a related language (e.g. Matlab, R) will be helpful. A Python tutorial will be provided and code fragments and hints will be provided on the homework assignments. Submission shall be via gradescope. After making a serious attempt to solve all homework problems on your own, students may consult with your classmates and friends and research ideas on the internet, but you must submit your own ideas and own work. Do not copy any text or code from your classmates, textbooks, or the internet – this is plagiarism.

Each assigned homework problem will be graded out of three points, with 3 = fully or essentially correct, 2 = mostly correct but with one or more notable errors, 1 = some effort was made but mostly incorrect, and 0 = not submitted or very little effort evident. If a problem has multiple parts, each part will be worth three points. Fractional scores may be given. Since this is a large class, depending on grader and GSI availability, not all problems (or parts of problems) may be graded. In this case, we will randomly select a few (parts of) problems to grade, and your grade will be based on these. Your performance on ungraded problems will not be a consideration in determining your homework grade.

Every homework will ask the students to submit at least one multiple choice question on the current subject matter along with its solution. These are graded only on whether both a question and answer were provided.

These questions with solutions will be posted in Perusall for discussion, typo identification, etc. All proposed questions will be visible to all other students and could be used for the quizzes. Simple questions are not discouraged; more important is that the question be correctly posed with a correct answer provided. If the question is too vague, posed incorrectly, has no correct answer, or is otherwise faulty it may be removed from shared materials.

Project: We will take part in the 2022 Fall ML Reproducibility challenge. This will be a group project that you will work on throughout the semester with 3-4 students per team. I recommend you choose your group to have people with experience in both programming and theory. I opened the "search for teammates" Piazza post for you to find groups.

Deadlines for project assignments (PA) are as follows:
PA 1 - Sept 30 - Select your paper and give justification, worth 2% of the total project grade.
PA 2 - Oct 28 - Submit a draft reproducibility report, worth 30% of the total project grade.
PA 3 - Nov 4 - Submit peer review for 3 other reproducibility reports, worth 6% of the total project grade.
PA 4 - Nov 29, Dec 1, Dec 6, Dec 8 - Present your work to the class, worth 10% of the total project grade. Attendance for at least three of these presentation days is mandatory and worth 2% of the total project grade.
PA 5 - Dec 11 - Final report due, worth 40% of the total project grade.
PA 6 - Dec 15 - Final report peer review due, worth 10% of the total project grade.

The goal of the annual Machine Learning reproducibility challenge "is to encourage the publishing and sharing of scientific results that are reliable and reproducible. In support of this, the objective of this challenge is to investigate reproducibility of papers accepted for publication at top conferences by inviting members of the community at large to select a paper, and verify the empirical results and claims in the paper by reproducing the computational experiments, either via a new implementation or using code/data or other information provided by the authors." (Quote from `https://paperswithcode.com/rc2022`.)

For the project in this class, each small group of 3-4 students will take part in the challenge. **For class, you may select any paper to reproduce from NeurIPS, ICML, ICLR, AAAI and IJCAI-ECAI, as well as the journals JMLR and TMLR.** If other major ML conferences are added, we will update the list. The challenge also includes ACL, EMNLP, CVPR, ECCV, ACM FAccT, SIGIR, and the journal TACL, which include papers without any machine learning topics. If your group chooses to select a paper from one of these venues, it should have a major machine learning component for you to reproduce and evaluate their claims. You should highlight the machine learning aspects in your initial justification when selecting these papers and note that your selection is subject to approval by the instructional staff. Note: NeurIPS authors will be notified of acceptance around September 14. ICML, ICLR, AAAI, and IJCAI-ECAI have already released their accepted papers.

As highlighted in the challenge description, "reproducing" a paper for the challenge involves investigating their *claims and conclusions* and evaluating whether those claims and conclusions are reproducible. That is likely to include reproducing their experimental results on both synthetic and real data. Claims often are made more broadly, yet in the paper they may be only illustrated on a few datasets, so reproducing their claims could involve applying their algorithm with different hyperparameters, to new datasets, and/or to synthetic data with realistic variations from the original paper. If computational efficiency is a focus of the original paper, students may want to time the algorithms using their own computing resources to see if they align with the author's numbers. If the algorithm has several components, (additional) ablation studies may be carried out. Claims may also involve mathematical statements that should be carefully checked, but simply saying "We checked the proof and it is correct" is not enough – one must look at parts of the proofs or derivations where steps are missing, flesh out those steps, include additional references for justification, etc. Interaction with the original authors is encouraged. Groups may use code that authors share online or by email. Students may also choose to implement the algorithms from scratch, especially if the paper claims that the algorithms are simple to implement. You are encouraged to submit your work to the challenge (deadline February) but you are not required to do so.

The grades for PA 2 and PA 5 will be based on peer reviews. Students will grade each others' projects according to a specified rubric and using a conference-style peer review system, overseen by the instructors. We expect that you take the peer reviews very seriously, and we will take care to ensure fairness.

Presentations will be short; the exact amount of time will depend on the total number of groups in the class. You are expected to get the high-level ideas of your project across, and all students in the group must present at least one slide (excluding the title slide). If you find that you must miss at least 2 presentation days due to illness or symptoms, please inform the entire teaching staff with a private Piazza message as soon as possible and before the second missed day with subject line "EECS 553 - presentation attendance". We will accommodate you.

Participation: There will be two main opportunities for participation: Piazza and iClicker. If you are regularly active and engaged using these formats, you will get a high score for participation.

Piazza is used as a message board – a place to ask and answer questions, share thoughts, discuss ideas, applications, and algorithms, etc. Students should post a question/comment, share a resource, or answer a question – and it must be visible to the whole class, though anonymously is fine – once per week on average to get full credit. I especially encourage the sharing/discussion of typos in the notes or other shared sources. Scores will be tallied monthly.

For iClicker, I will periodically ask questions during class to make sure everyone is following along. Each question is worth 3 points. Each question will (hopefully) have one best answer, but may have multiple answers that count as "right." You will get 2 points for attempting the question and 3 points for getting one of the right answers. The **ten days** of lowest scores will be dropped.

COVID-related policies: EECS 553 is guaranteed to be highly enrolled. We therefore must be extremely vigilant and thoughtful of our fellow classmates and our own health when attending class in person. Passing the UM health screen is required. Masks or other distancing policies will be implemented if reinstated by the university. All lectures will be recorded using the UM lecture capture service and available for viewing after class (asynchronous). The above drop policies should be sufficient in the case that you cannot make it to class because of sickness, exposure, or symptoms, however if you have exhausted your drop allowances and have symptoms please inform the entire teaching staff with a private Piazza message as soon as possible with details with subject line "EECS 553 - illness". Attendance is required for quizzes and final project presentations, but for the other class meetings there is no need to inform the instructional staff if you are sick / cannot attend. If the professor has symptoms, an announcement will be made through Canvas, and class will be on zoom with link in Canvas>Zoom.

Attendance: During COVID attendance requirements need to be flexible so that students don't come to class sick, with symptoms, or after being exposed. My policies are designed to be flexible. Nevertheless, please don't enroll for this class if you will be remote for much of the semester or have other classes or obligations during our class time. I will never make accommodations for a student who has double-booked themselves, and I promise it's not worth it to try to juggle two courses/obligations in the same time slot. iClicker is worth 5% of your grade and requires attendance, and attendance is required at quizzes (in the evenings) and presentation days (both with exceptions for sickness listed above).

Computing resources: Students may remotely connect to CAEN lab computers. These computers are intended for interactive use, as opposed to long-running, resource intensive jobs.
`https://caen.engin.umich.edu/connect/`

In addition, for the final project, each student will be granted a certain budget of computing time on the Great Lakes computing cluster, which includes both CPUs and GPUs. Use of the cluster is entirely optional. Students are advised that scheduling jobs, especially GPU jobs, requires careful setup and the 553 instructors cannot guarantee support for all student questions. In other words, the service is offered "as is."

That said, students may ask questions of the class through Piazza as well as the ARC-TS staff. Because of the time needed to set up jobs (again, especially GPU jobs), and delays in responses to getting technical questions answered, students should weigh the pros and cons of using the cluster. Starting early is definitely recommended. Difficulties using the cluster will not be a valid excuse for incomplete project experiments.

As an alternative to Great Lakes, students may also consider Google Colaboratory, which gives each student access to a free albeit slow GPU.

Honor policy: I take academic integrity very seriously. The College of Engineering is a community in which personal responsibility, honesty, fairness, respect, and mutual trust are maintained. You are expected to practice the highest possible standards of academic integrity. Any deviation from this expectation will result in a minimum academic penalty of your failing the assignment, and will result in additional disciplinary measures. This includes, but is not limited to, cheating, using unauthorized material during exams, using or copying another student's work, copying solutions from internet materials, plagiarism, and any other form of academic misrepresentation or misconduct. In particular, any 553 materials that are found on coursehero, chegg, etc will be immediately reported to the honor council.

For a list of actions that constitute misconduct, and possible sanctions for those actions, please see the Code of Conduct at `http://www.engin.umich.edu/college/academics/bulletin/rules`. I will not hesitate to enforce these standards in 553.

Commitment to equal opportunity: The Faculty of the COE are committed to a policy of equal opportunity for all persons and do not discriminate on the basis of race, color, national origin, age, marital status, sex, sexual orientation, gender identity, gender expression, disability, religion, height, weight, or veteran status. Please feel free to contact your instructors with any problem, concern, or suggestion. We ask that all students treat each other with respect.

Diversity, Equity, and Inclusion: I consider this classroom to be a place where you will be treated with respect, and I celebrate individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability, and other visible and nonvisible differences. All members of this class are expected to contribute to a respectful, welcoming and inclusive environment for every other member of the class. I am dedicated to helping each of you achieve all that you can in this class. I may, either in lecture or smaller interactions, accidentally use language that creates offense or discomfort. Should I do this, please contact me to help me understand and avoid making the same mistake again. If you are someone whose background is underrepresented in electrical engineering and computer science, I want to be your advocate.

Accommodations for Students with Disabilities: If you think you need an accommodation for a disability, please let you instructor know at your earliest convenience. Some aspects of this course may be modified to facilitate your participation and progress. As soon as you make us aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; `http://ssd.umich.edu`) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Student Mental Health and Wellbeing: University of Michigan is committed to advancing the mental health and wellbeing of its students, as am I. While I am not a mental health expert, I have personal experience with mental health struggles, and I support you in identifying these challenges and seeking help. Especially during COVID times it is critical that you monitor and care for your mental health. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and `https://caps.umich.edu/` during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (734) 764-8320 and `https://www.uhs.umich.edu/mentalhealthsvcs` or `https://uhs.umich.edu/stressresources`,

or for alcohol or drug concerns, see `www.uhs.umich.edu/aodresources`. For a listing of other mental health resources available on and off campus, visit:
`https://caps.umich.edu/article/um-mental-health-resources`.

Auditing: To audit the course, students should enroll as a visitor. In the interest of public health, we ask (but unless the classroom is regularly completely full, we will not enforce) that auditors do not attend class but rather watch the lecture recordings. There are no required assignments for auditors/visitors. Your program coordinator can help you enroll as a visitor. Please print this page to pdf and provide to your program coordinator as proof of our approval to enroll as a visitor. See also the Rackham policy for Visiting:
`https://rackham.umich.edu/academic-policies/section3/`

Override policy: This was our prior override policy, but as of September 15, 2022, the class is not full, so any students can join.

We will give preference to students who have taken the pre-requisite coursework of graduate probability and linear algebra at UM. If you haven't taken these courses, but have undergraduate level exposure to these concepts, I encourage you to consider EECS 498 - Principles of Machine Learning. The courses that I am currently considering as satisfying the requirement are (one from each category):

- Probability: EECS 501, Stats 510, Stats 610, Math/Stats 525, Math/Stats 625, Math/Stats 626, Biostats 601, Biostats 801, Biomede 503, ECON 503, SI 544
- Linear algebra/Matrix methods/Linear systems: EECS 505, EECS 551, Aerosp 550, CEE 571, MECHENG 564, EECS 560, ROB 501

We will admit students from the waiting list in the following order. Within each group, order is determined by position on the waiting list. I also list the number of such students currently on the waiting list, with the qualifier "potentially" since the waiting list does not identify whether students have taken the above courses.

- EECS PhD students who satisfy both graduate coursework requirements. (potentially 16 as of 7/5/2022)
- ECE MS students who satisfy both graduate coursework requirements. (potentially 105 as of 7/5/2022)
- ECE MEng Data Sci & Machine Learning students who satisfy both graduate coursework requirements. (potentially 21 as of 7/5/2022)
- IOE/Statistics/Math/AIM PhD students who satisfy both graduate coursework requirements. (potentially 6 as of 7/5/2022)
- EECS undergraduate students, other ECE MEng students, and CSE/Data Science MS students who satisfy both graduate coursework requirements. (potentially 28 as of 7/5/2022)
- Other PhD students who satisfy both graduate coursework requirements. (potentially 28 as of 7/5/2022)
- Other MS/MEng/undergraduate students who satisfy both graduate coursework requirements.

Course plan (subject to change):

Tues Aug 30 - Intro, Probabilistic Setup for Classification, Risk/Empirical Risk,
HW 1 released
Thurs Sept 1 - Unconstrained Optimization
Tues Sept 6 - Bayes Classifier, LDA, HW 1 on probability/matrices due, HW 2 released
Thurs Sept 8 - QDA, Naive Bayes
Tues Sept 13 - Logistic Regression, Linear Regression
Thurs Sept 15 - Empirical Risk Minimization, Regularization
Tues Sept 20 - Separating Hyperplanes, HW 2 due, HW 3 released
Thurs Sept 22 - Constrained optimization, Support Vector Machines
Tues Sept 27 - SVM cont'd, Kernel methods
Thurs Sept 29 - SGD, Multilayer perceptron, backpropagation
Fri Sept 30 - Selection of ML paper due
Tues Oct 4 - Neural networks topics cont'd, HW 3 due
Thurs Oct 6 - Catch up and/or review
Tues Oct 11 - Q&A in class, Quiz, HW 4 released
Thurs Oct 13 - Deep learning
== Fall Break ==
Thurs Oct 20 - Unsupervised learning introduction, Dimensionality reduction, PCA
Tues Oct 25 - Clustering, K-means, Gaussian Mixture Model, HW 4 due
Thurs Oct 27 - TBD
Fri Oct 28 - First draft ML report due
Tues Nov 1 - TBD
Thurs Nov 3 - Learning to Rank, Rank aggregation
Fri Nov 4 - Peer review of ML reports due
Tues Nov 8 - Matrix factorization, Matrix completion, HW 5 released, Election Day
Thurs Nov 10 - Nonlinear dimensionality reduction, Autoencoders, Embeddings
Tues Nov 15 - Anomaly detection, Distribution shift
Thurs Nov 17 - Fairness in Machine Learning
Tues Nov 22 - Open Q&A, HW 5 due
== Thanksgiving ==
Tues Nov 29 - Presentations
Thurs Dec 1 - Presentations
Tues Dec 6 - Presentations
Thurs Dec 8 - Presentations,
Concluding lecture about where the future might lead (20 mins)
== After classes end ==
Fri Dec 9 - Final ML report due
Thurs Dec 15 - Peer review of final ML reports due